Inverse Problems **15** (1999) 1469–1486. Printed in the UK PII: S0266-5611(99)05779-2

# A non-uniqueness problem in solar hard x-ray spectroscopy

M Piana†, R Barrett‡, J C Brown‡ and S W McIntosh‡§

† INFM, Dipartimento di Fisica, Università di Genova, via Dodecaneso 33, I-16146 Genova, Italy
‡ Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
§ High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO 80307-3000, USA

Received 7 July 1999

**Abstract.** We consider the hard x-ray emission process by interaction between the electrons and the ions in the solar atmosphere. We provide the integral equations describing this process as an inverse problem in the case of uniform ionization of the plasma and of a simple but rather realistic approximation of non-uniform conditions. The singular system of the integral operators is computed analytically in the continuous case for the uniform ionization model and numerically in the case of discrete data for both uniform and non-uniform ionization conditions. By analytical arguments and analysis of the singular spectrum we point out that non-uniform ionization results in an ambiguous interpretation of the solution of the integral equation, this solution not being unique. Finally, we briefly recall that this analysis facilitates methods for recovering unique and regularized solutions from high-resolution hard x-ray spectral data soon to be forthcoming from the HESSI space mission.

## 1. Introduction

In July 2000 NASA's Hessi mission [6] will acquire many high-resolution x-ray spectral data from the Sun. The impending launch of this mission has spurred interest in the problem of processing these data in order to infer information about the emission phenomena occurring in the solar atmosphere. In the present paper we show that this problem can be formulated as a linear inverse problem. We first consider a rather primitive model of uniform ionization for the solar atmosphere. Then we point out a fundamental ambiguity arising in the solution of the problem when the more realistic condition of non-uniform ionization is assumed. Finally, we propose a numerical analysis providing mathematical tools for a physically significant unique solution of the problem.

Cosmic plasmas such as solar active regions and compact stellar accretion discs are periodically characterized by transient excitation of the atmosphere followed by a dramatic increase of the electromagnetic emission and by the ejection of plasma particles outside the atmosphere. It is widely recognized that these explosive phenomena have magnetic origin, all other possible energy sources (such as gravitational and thermal sources) being insufficient to explain the amounts of energy released, of order $10^{32}$ erg in the case, for example, of a solar flare [15]. However, details of how energy release occurs and how a substantial fraction of it is channelled into accelerating plasma particles are poorly understood. As far as the energy release mechanism is concerned, it must be noticed that the equations of resistive magnetohydrodynamics (MHD), the theoretical basis for the description of this kind of process, involve a diffusion time which is much larger than the experimentally observed release time. In order to explain such a fast energy release many model situations have been formulated:

0266-5611/99/061469+18$30.00  

e.g., via decay of a 'sheet current' separating regions of oppositely directed magnetic field which reconnect; untwisting of the azimuthal magnetic field in a twisted loop by decay of a current along the loop; and diverse MHD instabilities. For rather complete reviews of solar and cosmic flares, see [11, 14, 16]. Each of these processes will generate complicated distributions of electric field—both macroscopic (induced fields) and microscopic (plasma waves)—and also of mass flow, which will accelerate electrons and ions by various mechanisms such as direct runaway in large-scale fields, stochastic wave particle processes, and shock wave interactions. In none of these cases is the theory well established and of course a general goal to validate the model is to infer from data information about the resulting fast particle distributions (spectral, spatial and temporal). To this end two possible approaches can be addressed. In a direct route distinctive aspects of the distributions are predicted from models for comparison with observed data while in an inverse approach the particle distributions are inferred non-parametrically from data. Because many of the physically important functions are not observable directly (image resolution is too small and photon not particle spectra are measured), much recent effort has gone into the inverse approach and particularly into problems of inferring the distribution of accelerated particles and of heated plasma from their electromagnetic spectra, as potential diagnostics of the basic physical processes involved [5, 7, 10, 12]. The inverse problem approach to this has now become especially relevant following major advances in the resolution available in high-energy photon spectrometry.

The Sun, by far the nearest star, provides the best opportunity we have for study of these problems. In particular, during the impulsive phase of a solar flare, strongly accelerated electrons collide with the ions of the background plasma (bremsstrahlung process) thus losing their energy by emitting x-ray radiation in the range 10–100 keV. The inverse problem we are interested in is to restore the electron distribution function from the measured x-ray data. This solar flare hard x-ray (HXR) diagnostic problem was first recognized as an integral inversion problem in [3] where for the non-relativistic Bethe–Heitler cross section the problem is reduced to Abel's equation and hence solved analytically. However, only in recent years, with the advent of Ge detectors, have HXR data approached the spectral resolution and precision needed for such inversions to be undertaken meaningfully. The crucial difficulty in this inverse approach is in the numerical instability which dramatically shows up in the analytical solution of the problem when the data are affected by experimental noise. This instability, a consequence of the ill-posedness of the problem, can be reduced by using *ad hoc* regularization techniques as done, for example, in [7, 12, 13, 17].

However, besides instability, an important complication in the spectral inverse problem was recognized in [4] in the case of the so-called thick-target model, where the electrons emit bremsstrahlung HXRs in Coulomb collisions with the dense lower solar atmosphere as they descend from a coronal acceleration site. Most thick-target modelling of data approximates the atmospheric target as fully ionized, following [3]. However, in [4] it is emphasized that the fall in target ionization across the solar transition zone (from hot corona to cool chromosphere) is accompanied by a drop in the electron collisional energy loss rate which enhances the HXR yield and distorts the HXR spectrum compared with the ionized case for a given electron injection spectrum. In [4] it is found by analytical arguments that, in the case of a (good) step function approximation to the ionization change, the corresponding inverse problem has no unique solution for the electron injection spectrum even when the physical positivity constraint is applied.

In section 2 of the present paper we describe the thick-target bremsstrahlung problem both in the case of uniform ionization and in the case of a step function approximation to the ionization change. As far as the uniform case is concerned, a singular value decomposition (SVD) analysis of the integral equation describing the bremsstrahlung process is performed in

section 3 both from an analytical and from a numerical point of view. In the case of non-uniform ionization, following [4] in section 4, we analytically point out the non-uniqueness of the solution of the inverse problem even for physically meaningful electron distribution functions. Then we compute the singular system of the integral operator with a step function integral kernel and relate the behaviour of the singular spectrum to this non-uniqueness property. Finally, in section 5, we notice that this SVD analysis provides a basis for a robust approach to HESSI HXR data inversion since both the generalized solution as a means to eliminating non-uniqueness and the most typical regularized solutions as a means to reducing the numerical instability due to the ill-posedness of the thick-target inverse problem can be expressed as linear combinations of the singular functions.

Finally, we observe that the study of the sensitivity of the singular spectrum to changes in the integral kernel of the operator describing the process represents a goal of rather general interest in solar/stellar physics where the most realistic models are typically formulated in terms of integral equations with different kernels allowing for different physical conditions.

## 2. The thick-target inverse problem

Following the analysis of [3, 4] we can relate the bremsstrahlung spectrum $J(\epsilon)$ (i.e. photons emitted per second per unit of photon energy $\epsilon$) to the injection spectrum $F_0(E_0)$ (electrons per second per unit of electron injection energy $E_0$) by

$$H(\epsilon) = \frac{\epsilon K_0}{Q_0} J(\epsilon) = \int_\epsilon^\infty F_0(E_0) \int_\epsilon^{E_0} \frac{q(\epsilon, E)\,\mathrm{d}E\,\mathrm{d}E_0}{\lambda + x(E_0^2 - E^2)}. \tag{2.1}$$

Here $Q_0 q(\epsilon, E)$ is the bremsstrahlung cross section where $Q_0$ is a constant. The ionization $x$ along the electron path is a function of the column depth $N$ reached by the electron; $N$ can be expressed in terms of the electron energy $E_0$ at the injection and the electron energy $E$ at the point where the electron is decelerated by collisions. The constants $K_0$ and $\lambda$ are defined by $K = 2\pi e^4 \Lambda$ and $\lambda = \Lambda_{\mathrm{eH}}/\Lambda$ where $\Lambda = \Lambda_{\mathrm{ee}} - \Lambda_{\mathrm{eH}}$ and $e$ is the electron charge. $\Lambda_{\mathrm{ee}}$ and $\Lambda_{\mathrm{eH}}$ are the (constant) Coulomb logarithmics for the electron–electron and the electron–hydrogen collisions. In (2.1) the outer integral is over electron injection energy and the inner over the electron path through the target. In [9] many possible forms for the bremsstrahlung cross section $Q_0 q(\epsilon, E)$ can be found.

**Remark 2.1.** Here we are mainly interested in the relative effect of having the ionization $x$ varying as against being constant and, therefore, we consider the simplest Kramers form $q = 1$. In [12] the bremsstrahlung problem is studied in the case of uniform ionization and Bethe–Heitler cross section. However, the sensitivity to changing the cross section in a non-uniform ionization plasma, also allowing for relativistic effects, will be the subject of future research.

In [7, 12, 13] the approximation $x = 1$ is adopted. In that case by reversing the integration order and by differentiating with respect to $\epsilon$ equation (2.1) becomes

$$L(\epsilon) := -H'(\epsilon) = \frac{1}{\lambda + 1} \int_\epsilon^\infty F_0(E_0)\,\mathrm{d}E_0. \tag{2.2}$$

In [4] (2.1) is briefly discussed for general forms of $x$ but then interest is concentrated on the case where $x(N)$ is a step function from 1 to 0 at some depth $N_1$ to approximate well the actual solar atmospheric structure. If $E_1$ is the value of $E_0$ needed to reach $N_1$ this amounts to setting

$$x(E_0^2 - E^2) = \begin{cases} 1 & E_0^2 - E^2 \geqslant E_1^2 \\ 0 & E_0^2 - E^2 < E_1^2. \end{cases} \tag{2.3}$$

In this case for $q = 1$, reversing the integration order and differentiating with respect to $\epsilon$, (2.1) can be written as

$$-\int_{\epsilon}^{\infty} \frac{F_0(E_0)\,\mathrm{d}E_0}{\lambda + x(E_0^2 - \epsilon^2)} = -H'(\epsilon) = L(\epsilon) \tag{2.4}$$

for general $x$, while for $x$ given by (2.3) it becomes

$$\int_0^{\infty} F_0(E_0) K(\epsilon, E_0)\,\mathrm{d}E_0 = L(\epsilon) \tag{2.5}$$

where the kernel $K$ is

$$K(\epsilon, E_0) = \begin{cases} 0 & E_0 < \epsilon \\ \dfrac{1}{\lambda + 1} & \epsilon \leqslant E_0 \leqslant \sqrt{E_1^2 + \epsilon^2} \\ \dfrac{1}{\lambda} & E_0 > \sqrt{E_1^2 + \epsilon^2}. \end{cases} \tag{2.6}$$

In the following we will assume different variables for equations (2.2) and (2.5), (2.6). By defining $\eta = E_0^2/E_1^2$, $\xi = \epsilon^2/E_1^2$, $f(\eta)\,\mathrm{d}\eta = F_0(E_0)\,\mathrm{d}E_0/F_0(E_1)$, $g(\epsilon) = L(\epsilon)/L(E_1)$, equation (2.2) becomes

$$g(\xi) = \frac{1}{1+\lambda} \int_{\xi}^{\infty} f(\eta)\,\mathrm{d}\eta \tag{2.7}$$

while equations (2.5), (2.6) assume the neater form

$$g(\xi) = \frac{1}{\lambda + 1} \int_{\xi}^{\xi+1} f(\eta)\,\mathrm{d}\eta + \frac{1}{\lambda} \int_{\xi+1}^{\infty} f(\eta)\,\mathrm{d}\eta. \tag{2.8}$$

We finally notice that, in principle, the electron injection energy $E_0$ may assume whatever large value though it is reasonable to fix a maximum value significantly larger than the characteristic energy $E_1$. In other terms in the next sections we will consider and compare analysis of the equations

$$g(\xi) = \frac{1}{1+\lambda} \int_{\xi}^{c} f(\eta)\,\mathrm{d}\eta \tag{2.9}$$

and

$$g(\xi) = \frac{1}{\lambda + 1} \int_{\xi}^{\xi+1} f(\eta)\,\mathrm{d}\eta + \frac{1}{\lambda} \int_{\xi+1}^{c} f(\eta)\,\mathrm{d}\eta. \tag{2.10}$$

with $1 < c < \infty$.

## 3. SVD analysis (uniform ionization case)

In the present section we consider the study of equation (2.9) by means of a singular value decomposition analysis. We introduce the integral operator $S : L^2(0, c) \to L^2(0, c)$,

$$(Sf)(\xi) = \frac{1}{\lambda + 1} \int_{\xi}^{c} f(\eta)\,\mathrm{d}\eta \tag{3.1}$$

with $L^2(0, c)$ equipped with the canonical scalar product $(\cdot, \cdot)_2$

$$(f_1, f_2)_2 = \int_0^c f_1(\eta) f_2(\eta)\,\mathrm{d}\eta. \tag{3.2}$$

$S$ is a compact linear operator whose adjoint $S^* : L^2(0, c) \to L^2(0, c)$ is

$$(S^* g)(\eta) = \frac{1}{\lambda + 1} \int_0^{\eta} g(\xi)\,\mathrm{d}\xi. \tag{3.3}$$

The singular system of $S$ is defined as the infinite set of triples $\{\sigma_k; u_k, v_k\}_{k=1}^{\infty}$ such that

$$Su_k = \sigma_k v_k; \qquad S^* v_k = \sigma_k u_k. \tag{3.4}$$

The following theorem holds.

**Theorem 3.1.** *The singular system of $S$ is given by*

$$u_k(\eta) = a_k \sin\left(\frac{\eta}{\sigma_k(1+\lambda)}\right), \tag{3.5}$$

$$v_k(\xi) = b_k \sin\left(\frac{\xi - c}{\sigma_k(1+\lambda)}\right) \tag{3.6}$$

*and*

$$\sigma_k = \frac{2c}{\pi(1+\lambda)(2k-1)} \tag{3.7}$$

*with $b_k = (-1)^{k+1} a_k$ and $a_k = \sqrt{2/c}$.*

**Proof.** Equations (3.1)–(3.4) imply that

$$u_k(0) = 0; \qquad v_k(c) = 0. \tag{3.8}$$

Furthermore, by differentiating both sides of (3.4) and by using (3.8) we obtain

$$u_k'(c) = 0; \qquad v_k'(0) = 0. \tag{3.9}$$

Finally, from

$$(S^* S u_k)(\eta) = \frac{1}{(\lambda+1)^2} \int_0^{\eta} \int_{\xi}^{c} u_k(\eta)\, \mathrm{d}\eta\, \mathrm{d}\xi, \tag{3.10}$$

$$(S S^* v_k)(\xi) = \frac{1}{(\lambda+1)^2} \int_{\xi}^{c} \int_0^{\eta} v_k(\xi)\, \mathrm{d}\xi\, \mathrm{d}\eta \tag{3.11}$$

and equations (3.4), (3.8) and (3.9), we have that the singular functions $u_k$ and $v_k$ and the singular values $\sigma_k$ satisfy the boundary-value problems

$$\begin{aligned} &\sigma_k^2 u_k'' = -\frac{1}{(1+\lambda)^2} u_k \\ &u_k(0) = 0 \\ &u_k'(c) = 0 \end{aligned} \tag{3.12}$$

and

$$\begin{aligned} &\sigma_k^2 v_k'' = -\frac{1}{(1+\lambda)^2} v_k \\ &v_k(c) = 0 \\ &v_k'(0) = 0. \end{aligned} \tag{3.13}$$

These two problems can be easily solved to obtain equations (3.5), (3.6) and the boundary conditions imply equation (3.7) for the singular values. As far as the constants $a_k$ and $b_k$ are concerned, again differentiation of equation (3.4) implies that $b_k = (-1)^{k+1} a_k$ and the $a_k$ can be determined by imposing the normalization condition $\|u_k\|^2_{L^2(0,c)} = 1$ so that $a_k = \sqrt{2/c}$. □

$S$ represents an example of a compact operator which is meaningful from a physical point of view and whose singular system can be exactly computed. However, it must be noted that in real applications detectors provide the photon spectrum only in correspondence to a finite number of discrete photon energies (to be more precise, a finite number of integrals of the

photon spectrum over finite ranges of the photon energy; nevertheless here we choose to neglect this integration since the effects on the singular values do not seem to be significant [12]). In other terms it is completely realistic to consider the linear inverse problem with discrete data

$$g_n = S_N f \qquad n = 1, \ldots, N \tag{3.14}$$

with $S_N : L^2(0, c) \to Y$ the *single-step* finite rank operator

$$(S_N f)(\xi_n) = \int_{\xi_n}^{c} f(\eta)\, \mathrm{d}\eta \qquad n = 1, \ldots, N, \tag{3.15}$$

$g_n = g(\xi_n)$ $n = 1, \ldots, N$ and $Y$ is an Euclidean space equipped with the scalar product

$$(\boldsymbol{g}, \boldsymbol{h})_Y = \sum_{n,m=1}^{N} g_n w_{nm} h_m^* \tag{3.16}$$

$\forall \boldsymbol{g}, \boldsymbol{h} \in Y$. $S_N$ is again a compact linear operator and the singular system can be introduced as the finite set of triples $\{\sigma_k^{(N)}; u_k^{(N)}, \boldsymbol{v}_k^{(N)}\}_{k=1}^{N}$ such that

$$S_N u_k^{(N)} = \sigma_k^{(N)} \boldsymbol{v}_k^{(N)}; \qquad S_N^* \boldsymbol{v}_k^{(N)} = \sigma_k^{(N)} u_k^{(N)} \tag{3.17}$$

where $S_N^* : Y \to L^2(0, c)$ is the adjoint operator

$$(S_N^* g)(\eta) = \sum_{n,m=1}^{N} \chi_{[\xi_m, c]} g_n w_{nm} \tag{3.18}$$

with $\chi_{[\xi_m, c]}$ the characteristic function of the interval $[\xi_m, c]$. The problem of performing the SVD of $S_N$ can be reduced to a diagonalization problem in $Y$. First, we observe that the singular values $\sigma_k^{(N)}$ and the singular vectors $\boldsymbol{v}_k^{(N)}$ are the eigenvalues and the eigenvectors of the matrix $S_N S_N^*$. On the other hand, Riesz's lemma implies that the application of $S_N$ on $f$ can be represented as

$$(S_N f)_n = (f, \psi_n)_2 \tag{3.19}$$

where $\psi_n(\eta)$ is defined by

$$\psi_n(\eta) = \begin{cases} 0 & \eta \leqslant \xi_n \\ \dfrac{1}{1+\lambda} & \xi_n < \eta \leqslant c. \end{cases} \tag{3.20}$$

It can be shown straightforwardly [1] that $S_N S_N^* = \mathcal{G}^{\perp} W$ where $\mathcal{G}$ is the symmetric Gram matrix whose $nm$-entry is defined as $\mathcal{G}_{nm} = (\psi_n, \psi_m)_2$ and $W$ is the weight matrix,

$$W_{nm} = w_{nm}. \tag{3.21}$$

Moreover, the singular functions can be computed in terms of the $\psi_n$ by the equation

$$u_k^{(N)}(\eta) = \frac{1}{\sigma_k} \sum_{n,m=1}^{N} w_{nm} (\boldsymbol{v}_k)_m \psi_n(\eta). \tag{3.22}$$

Therefore, the crucial point is to determine the Gram matrix $\mathcal{G}$. In our case this computation is very easy and the result is, for $n \geqslant m$,

$$\mathcal{G}_{nm} = \frac{1}{(1+\lambda)^2}(c - \xi_n). \tag{3.23}$$

We have diagonalized $\mathcal{G}^{\perp} W$ in both cases of uniform and geometrical sampling. For uniformly sampled points

$$\xi_n = \xi_1 + d(n-1) \qquad n = 1, \ldots, N \tag{3.24}$$

the typical form for the weights is

$$w_{nm} = \mathrm{d}\delta_{nm} \tag{3.25}$$

while for the geometrical distribution

$$\xi_n = \xi_1 \Delta^{n-1} \qquad n = 1, \ldots, N \tag{3.26}$$

a good choice is

$$w_{nm} = (\log \Delta)\xi_n \delta_{nm}. \tag{3.27}$$

We compare the singular values of $S$ and $S_N$ in the case of $N = 100$ points in the energy range $\xi_1 = 0$, $\xi_N = 70$. First of all, we observe that if the uniform sampling is adopted the singular spectrum of $S_N$ mimicks the singular spectrum of $S$ in a more accurate way than if the geometrical sampling is adopted. For example, for $c = 70$, the rms error defined by

$$rms = \sqrt{\frac{\sum_{k=1}^{N}(\sigma_k - \sigma_k^{(N)})^2}{\sum_{k=1}^{N}\sigma_k^2}} \tag{3.28}$$

is 1.5% in the case of uniform sampling and 9.5% in the case of geometrical sampling. Therefore, in the following computations, uniform sampling will be adopted. Furthermore, as shown in table 1, the rms error is rather sensitive to the choice of the truncation parameter $c$. In fact, for $c = \xi_N = 70$, one row of the Gram matrix (3.23) vanishes and correspondingly $\sigma_N^{(N)}$ is also equal to zero. This implies that the discrepancy between the analytical and the discrete data singular spectra first decreases (until the Gram matrix has a row close to zero) and then increases since for increasing values of $c$ a fixed value of sampling points $N$ must cover an increasing energy range, thus reducing the accuracy of the computation. In figure 1 we compare the analytical singular spectrum with the numerical one obtained by using uniform sampling. In figure 2 the first four singular functions of $S_N$ (computed by equation (3.22)) are plotted together with the first four singular functions of $S$ (computed by equation (3.5)). From these results it follows that the data discretization does not significantly modify the singular system if a reasonably large number of uniformly sampled points is used.

**Table 1.** Comparison between the singular spectra of $S$ and $S_N$ with $N = 100$ uniformly sampled points in the range $\xi_1 = 0$, $\xi_N = 70$ for increasing values of the truncation parameter $c$.

| $c$ | rms error |
|---|---|
| 70 | $1.50 \times 10^{-2}$ |
| 71 | $1.40 \times 10^{-2}$ |
| 72 | $1.30 \times 10^{-2}$ |
| 73 | $1.21 \times 10^{-2}$ |
| 74 | $1.14 \times 10^{-2}$ |
| 75 | $1.10 \times 10^{-2}$ |
| 76 | $1.11 \times 10^{-2}$ |
| 77 | $1.17 \times 10^{-2}$ |
| 78 | $1.29 \times 10^{-2}$ |
| 79 | $1.45 \times 10^{-2}$ |
| 80 | $1.65 \times 10^{-2}$ |
| 85 | $2.91 \times 10^{-2}$ |

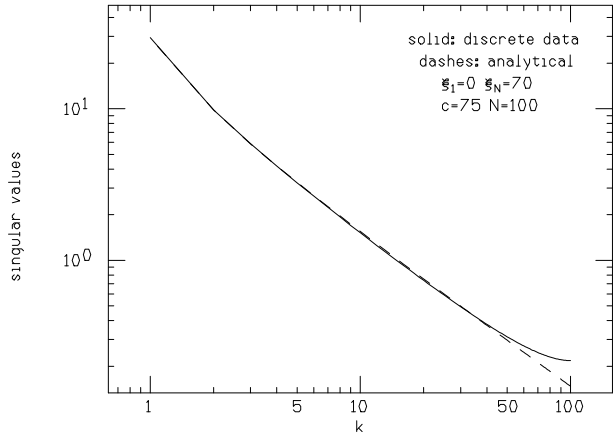


**Figure 1.** Singular values in the discrete data case (solid) and in the analytical case (dashed) for $N = 100$ uniformly sampled observation points in the range $\xi_1 = 0$, $\xi_N = 70$. The truncation parameter is $c = 75$.

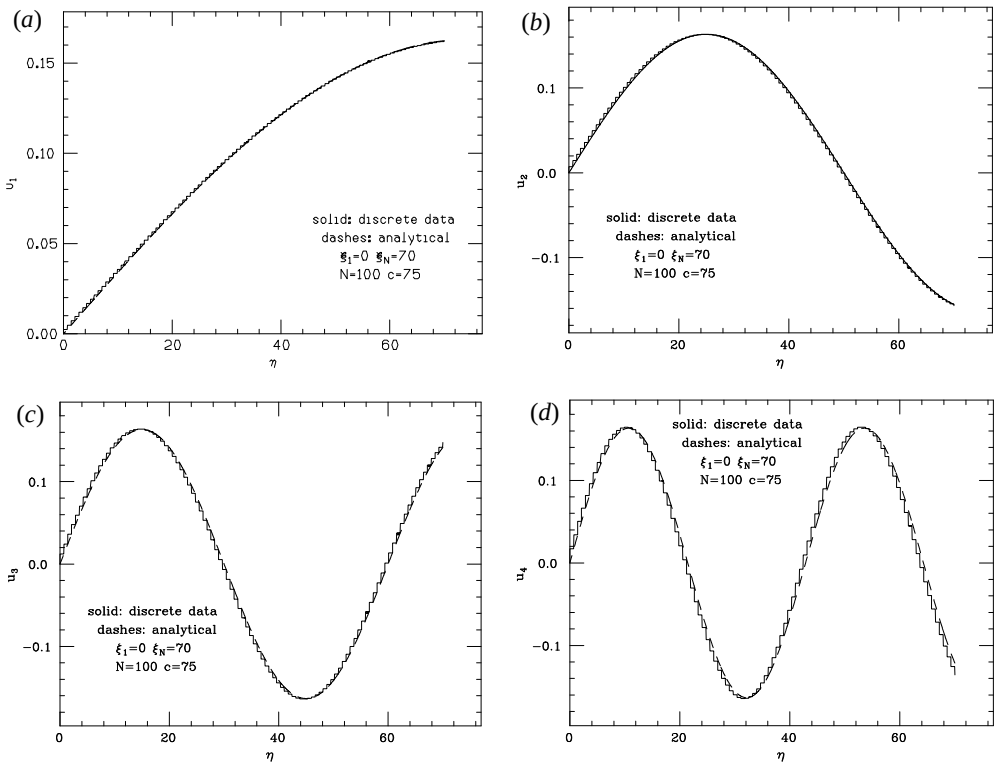


**Figure 2.** Singular functions $u_k$ in the discrete data case and uniform sampling (solid) and in the continuous data case (dashed): (*a*) first singular function $u_1$; (*b*) second singular function $u_2$; (*c*) third singular function $u_3$; (*d*) fourth singular function $u_4$. The parameters $N$, $\xi_1$, $\xi_N$ and $c$ are as in figure 1.

## 4. The non-uniform ionization case

Though widely used in data analysis for simplicity, a uniformly ionized plasma does not represent a realistical model for the description of thick-target bremsstrahlung in the solar corona. However, the effects of non-uniform ionization may be dramatic and lead to ambiguous interpretation. A first generalization is given by a step-function ionization structure with depth mimicking the sharp coronal–chromospheric separation in solar flare plasmas. If this model is adopted, the integral equation relating the injection spectrum and the photon radiation is given by equation (2.8), i.e. the presence of a non-uniform ionization is reflected in a double-step form of the integral kernel. This apparently simple and in some sense natural modification implies serious consequences for the study of the inverse problem with particular reference to the uniqueness of the solution.

### *4.1. Non-uniqueness: an analytical approach*

Many of the more surprising properties of the integral equation (2.8) can be appreciated through a simple analytic approach. Following [4], we differentiate (2.8) with respect to $\xi$ (making the natural assumptions that $g(\xi)$ is suitably differentiable and that the electron spectrum falls off sufficiently fast so that the total energy is finite, so $f \to 0$ as $\xi \to \infty$):

$$f(\xi) + \frac{1}{\lambda} f(\xi + 1) = f^*(\xi), \tag{4.1}$$

where $f^*(\xi) = -(\lambda + 1)g'(\xi)$ (this apparently odd notational choice is motivated by the fact that, as is obvious from (2.7), $f^*(\xi)$ is precisely the solution of the integral equation for *uniform* ionization, so that the non-uniform ionization problem can be viewed as a two-stage inversion, where the photon spectrum is first inverted assuming uniform ionization to obtain $f^*(\xi)$ and then this is used to solve for the *real* electron spectrum via (2.7), as described in the appendix). Equation (4.1) is not a solution for $f$ but is essentially a recurrence relation relating the solution at values of $\xi$ differing by one (which corresponds in physical units to $E_1$, the energy required to reach the transition region). With this interpretation it is clear that the values of the solution at points not precisely separated by an integer are unrelated (although often one would wish to impose some regularity condition, such as continuity or even analyticity, that would constrain the differences between neighbouring points). Moreover, on the face of it we are free to specify the solution arbitrarily in any (half-open) interval $(\xi, \xi + 1]$ and use (4.1) recursively to obtain $f(\xi)$ at any other point. Since we are only interested in positive $\xi$ we will choose $(0, 1]$ to be this canonical interval in what follows and define, for any $\xi$

$$\xi = i + \xi_0 \tag{4.2}$$

where $0 < \xi_0 \leqslant 1$ and $i$ is the largest integer less than $\xi$. Then for any function $\phi(\xi_0)$ a solution of (4.1) is given by

$$f(\xi) = f(i + \xi_0) = (-\lambda)^i \left[ \phi(\xi_0) - \sum_{j=0}^{i-1} (-\lambda)^{-j} f^*(\xi_0 + j) \right]. \tag{4.3}$$

That this is true for *any* $\phi$ indicates the severity of the non-uniqueness.

In general, the solutions (4.3) will not be continuous (or will have discontinuous derivatives) at integer values of $\xi$, and it might be thought that the non-uniqueness is a result of allowing this lack of regularity, so that requiring, say, infinitely differentiable or analytic solutions might eliminate the non-uniqueness. That this is not so can easily be seen from an investigation of the null functions of (4.1) (i.e., the solutions of (4.1) with $f^*(\xi) \equiv 0$). The functions

$$f(\xi) = \lambda^{\xi} \sin(2n + 1)\pi\xi, \tag{4.4}$$

for any integer $n$, are obviously analytic and can be seen to satisfy (4.1) with $f^*(\xi) = 0$ (it is also possible to include an arbitrary phase in (4.4)). Thus, if an analytic solution to (4.1) exists, there will be infinitely many such solutions differing by linear combinations of the null functions in (4.4). In fact, given a $\phi(\xi_0)$ defined on $(0, 1]$ it is possible to expand $\lambda^{-\xi_0}\phi(\xi_0)$ as a Fourier series, so that more general null functions can be expressed in terms of those in (4.4).

The preceding analytic results on non-uniqueness overlook the important physical constraint that the real electron spectrum must be non-negative everywhere. Since general null functions, which from (4.3) are given by

$$f(\xi) = (-\lambda)^i \phi(\xi_0), \tag{4.5}$$

must alternate in sign from one integer-length interval to the next they cannot be positive definite, which might suggest that enforcing positivity eliminates the non-uniqueness. Again, as was shown in [4], this is not the case. Returning to (4.3) it is clear that imposing $f(\xi_0+i) \geqslant 0$ for varying $i$ results alternately in upper and lower bounds on the (otherwise arbitrary) function $\phi(\xi_0)$ (owing to the $(-\lambda)^i$ factor). So, for any positive integer $k$, we must have

$$\sum_{j=0}^{2k-1} (-\lambda)^{-j} f^*(\xi_0 + j) \leqslant \phi(\xi_0) \leqslant \sum_{j=0}^{2k} (-\lambda)^{-j} f^*(\xi_0 + j). \tag{4.6}$$

Clearly, therefore, the values of $\phi$ at any $\xi_0$, and hence of $f$ at any $\xi$, are fixed uniquely whenever the series in (4.6) converges absolutely, i.e. if

$$S(\xi_0) \equiv \sum_{j=0}^{\infty} |(-\lambda)^{-j} f^*(\xi_0 + j)| \tag{4.7}$$

converges, for then the upper and lower bounds must be equal, and the *unique* solution is given from

$$\phi(\xi_0) = \sum_{j=0}^{\infty} (-\lambda)^{-j} f^*(\xi_0 + j), \tag{4.8}$$

which gives $f$ via (4.3). Note that convergence of (4.7) is a sufficient but not a necessary condition for the solution to be unique: if it happens that the upper and lower bounds are equal at some finite $k$ (because $f^*(\xi)$ is zero over an interval of length one or greater, for example), then uniqueness is assured regardless of the behaviour of $S(\xi_0)$.

$S(\xi_0)$ will converge whenever $|\lambda| > 1$ (since the finite-energy condition ensures that $f^*(\xi)$ tends to zero for large $\xi$). For $|\lambda| < 1$, $S(\xi_0)$ only converges if $f^*$ falls off sufficiently fast to counteract the increase in $|\lambda|$, i.e.

$$|\lambda^{-j} f^*(\xi_0 + j)| = \mathrm{O}(j^{-1}). \tag{4.9}$$

As noted in [4], a sufficient (although again not necessary) condition for this is $f^*(\xi) = \mathrm{O}(\mathrm{e}^{-a\xi})$, with $a > -\log\lambda$. The common assumption of a power-law photon spectrum (extending to infinite energies) does not satisfy this, and non-uniqueness results.

Obviously, a real photon spectrum will not extend to arbitrary energies, and the preceding discussion shows that if the photon spectrum is flat ($f^*(\xi) = 0$) over any $\xi$-range of length one or larger the solution will be unique, so that a real photon spectrum will always give rise to a unique solution. However, the solution obtained will be highly sensitive to the chosen upper cut-off. From (4.8) the solution for a photon spectrum with an upper cut-off at $\xi = i_0$, say, is

$$\phi_{i_0}(\xi_0) = \sum_{j=0}^{i_0-1} (-\lambda)^{-j} f^*(\xi_0 + j). \tag{4.10}$$

If the cut-off was instead at a slightly larger $\xi = i_0 + \delta$, then for $\xi_0 < \delta$ the solution is

$$\phi_{i_0+\delta}(\xi_0) = \phi_{i_0} + (-\lambda)^{-i_0} f^*(\xi_0 + i_0), \tag{4.11}$$

and the last term will, in general, be large, for small $\lambda$ and/or large $i_0$. Thus, in practical situations the non-uniqueness is replaced by an instability to the form of the photon spectrum at large energies. This is particularly relevant to the results presented in this paper because we are able to study only the truncated problem numerically.

### *4.2. SVD analysis*

As in section 2, we assume that the electron energy is truncated at a sufficiently large value and that the emitted spectrum is detected in correspondence with $N$ discrete values of the photon energy uniformly sampled over a finite range. Therefore, it is natural to introduce the operator $D_N : L^2(0, \infty) \to Y$ such that

$$(D_N f)(\xi_n) = \frac{1}{\lambda + 1} \int_{\xi_n}^{\xi_n+1} f(\eta)\, \mathrm{d}\eta + \frac{1}{\lambda} \int_{\xi_n+1}^{c} f(\eta)\, \mathrm{d}\eta \qquad n = 1, \ldots, N \tag{4.12}$$

where $c$ is the usual truncation parameter. This allows us to write the inverse problem of determining $f$ from $g_n = g(\xi_n)$, $n = 1, \ldots, N$ in the synthetic form

$$g_n = D_N f. \tag{4.13}$$

$D_N$ is again a compact linear operator and the explicit form of the adjoint $D_N^* : Y \to L^2(0, c)$ is

$$D_N^* \boldsymbol{g}(\eta) = \sum_{n,m}^{N} \left\{ \frac{g_n}{1+\lambda} w_{nm} \chi_{[\xi_m, \xi_m+1]}(\eta) + \frac{g_n}{\lambda} w_{nm} \chi_{[\xi_m+1, c]} \right\}. \tag{4.14}$$

As in the case of the single-step operator $S_N$, $D_N$ can be represented in the form

$$(D_N f)_n = (f, \psi_n)_2 \tag{4.15}$$

where this time the functions $\psi_n$ are defined as

$$\psi_n(\eta) = \begin{cases} 0 & \eta \leqslant \xi_n \\ \dfrac{1}{\lambda + 1} & \xi_n < \eta < \xi_n + 1 \\ \dfrac{1}{\lambda} & \xi_n + 1 \leqslant \eta \leqslant c. \end{cases} \tag{4.16}$$

The singular system is as usual defined by the *shifted* eigenvalue problem

$$D_N u_k = \sigma_k v_k; \qquad D_N^* v_k = \sigma_k u_k \qquad k = 1, \ldots, N \tag{4.17}$$

and can be computed by diagonalizing the matrix $D_N D_N^*$. Relation $D_N D_N^* = \mathcal{G}^{\perp} W$ holds also in this case and the explicit form for the $nm$ entry of the Gram matrix is, for $n \geqslant m$,

$$\mathcal{G}_{nm} = \frac{1}{\lambda(\lambda + 1)} + \frac{1}{\lambda^2}(c - \xi_n - 1) \tag{4.18}$$

if $\xi_m + 1 \leqslant \xi_n$ and

$$\mathcal{G}_{nm} = \frac{1}{(\lambda + 1)^2}(\xi_m + 1 - \xi_n) + \frac{1}{\lambda(\lambda + 1)}(\xi_{n+1} - \xi_{m+1}) + \frac{1}{\lambda^2}(c - \xi_n - 1) \tag{4.19}$$

if $\xi_m + 1 > \xi_n$.

In figure 3 we plot the singular spectrum of $D_N$ and $S_N$ in the case of $N = 100$ points uniformly sampled between $\xi = 0$ and $\xi = 70$. The main difference between these two spectra is in the fact that the singular spectrum of $D_N$ presents a flat region in correspondence with

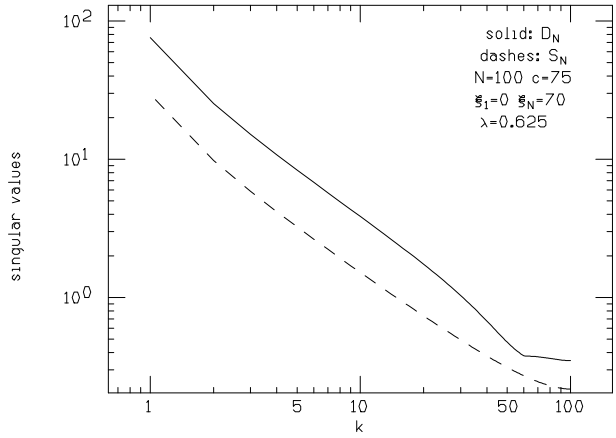


**Figure 3.** Comparison of the singular values of the double-step operator (solid) with the singular values of the single-step operator (dashed). The parameters $N$, $\xi_1$, $\xi_N$ and $c$ are as in figure 1 and $\lambda = 0.625$.

high values of $k$. In order to interpret this behaviour, we introduce the linear integral operator $D$,

$$(Df)(\xi) = \frac{1}{\lambda + 1} \int_{\xi}^{\xi+1} f(\eta)\, \mathrm{d}\eta + \frac{1}{\lambda} \int_{\xi+1}^{\infty} f(\eta)\, \mathrm{d}\eta. \tag{4.20}$$

As explained in section 4.1, this operator has a non-trivial kernel, i.e. there exist non-zero functions $\{f\}$ such that $Df = 0$. These functions are also in the kernel of $D^*D$ with $D^*$ the formal adjoint of $D$ and therefore the singular value $\sigma = 0$ has multiplicity bigger than one. Now the finite-rank operator $D_N$ can be considered as a truncated discrete-data approximation of $D$. Therefore, in particular, the high-order singular values of $D_N$ (i.e. the smallest ones), which are indeed very close, can be seen as the truncated discrete-data approximation of the singular value zero of $D$ with multiplicity bigger than one. Of course, $D_N$ approximates $D$ the better, the larger the values of the truncation parameter $c$ and the number of sampling points $N$. For example, in figure 4 we fix $\lambda = 0.625$ (this value of $\lambda$ can be physically motivated with arguments from solar plasma physics) and consider different values of $c$ and $N$ (for physical reasons, once $c$ is fixed, the data variable range $(\xi_1, \xi_N)$ must be below $c$); as you can see the flat region in the singular spectrum is present only when the values of $c$ and $N$ provide a good trade-off between a large energy range and an accurate sampling. Indeed, for fixed $N$, the *plateau* disappears if $c$ is too small (figure 4(*a*)) and appears again for a larger value of $c$ (figure 4(*b*)). An analogous behaviour is assumed by the spectrum when $c$ is fixed and $N$ increases. For example, in figure 4(*c*) there is no *plateau* owing to undersampling while the flat region can be distinguished again in figure 4(*d*) where $N$ is sufficiently large to assure an accurate approximation of the singular spectrum of $D$. We also point out that the presence of a degenerate small singular value is related to the presence of a pronounced double-step behaviour in the integral kernel. In fact, in figure 5, it is shown that the flat region for high $k$ occurs only for values of $\lambda$ such that this double-step feature is significant (figures 5(*b*) and (*c*)) while if the difference between the two steps is small, i.e. for values of $\lambda$ too small or too large, no flat region can be observed (figures 5(*a*) and (*d*)). What this means physically is that in these two limits the bremsstrahlung spectrum becomes dominated by one of the uniform ionization limits $x = 1$ or $x = 0$. The first is that of equation (2.9) while the second is the same with $g$ rescaled by the factor $\lambda + 1$, both having analytically unique solutions.

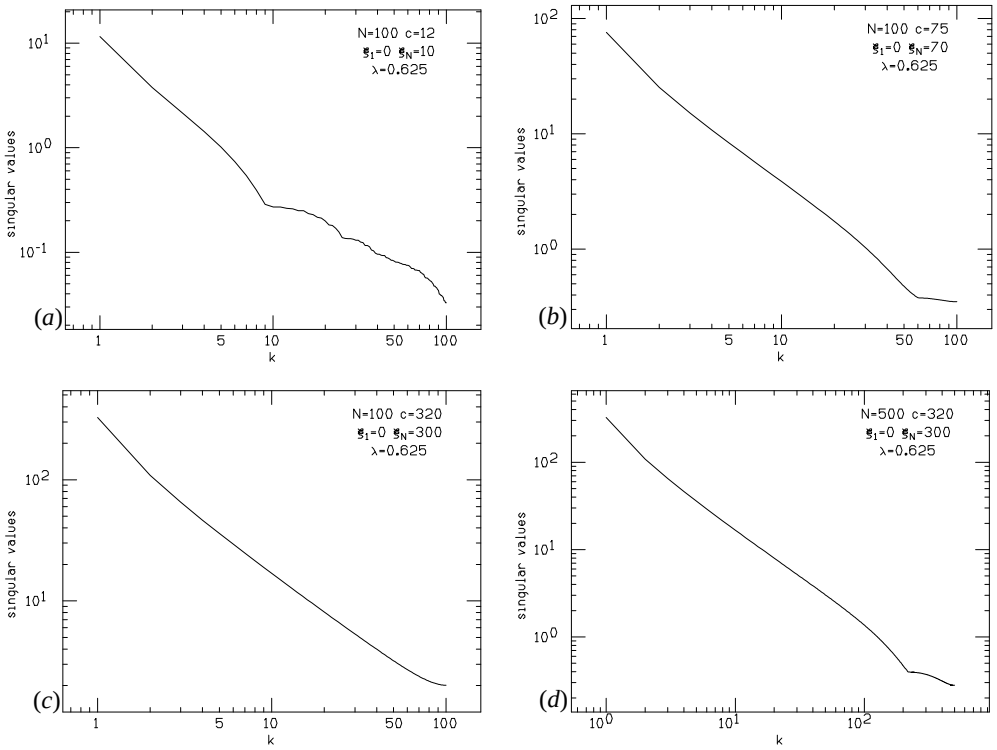


**Figure 4.** Singular spectrum of $D_N$ for different values of the truncation parameter and of the number of sampling points. (*a*) $\xi_1 = 0$, $\xi_N = 10$, $c = 12$; (*b*) $\xi_1 = 0$, $\xi_N = 70$, $c = 75$; (*c*) $\xi_1 = 0$, $\xi_N = 300$, $c = 320$, $N = 100$; (*d*) $\xi_1 = 0$, $\xi_N = 300$, $c = 320$, $N = 500$. For each plot $\lambda = 0.625$.

In figures 6 and 7 some singular functions of $D_N$ are plotted in the case where $N = 100$, $\xi_1 = 0$, $\xi_N = 70$, $c = 75$, $\lambda = 0.625$ (the same conditions as figure 4(*b*)). In particular, in figure 6 we present singular functions corresponding to singular values in the steep part of the spectrum. These functions oscillate with frequency increasing with their order, as typically happens to this kind of integral operator [2]. However, for singular functions corresponding to the flat region of the spectrum, such as the ones represented in figure 7, a low-frequency oscillation behaviour can be noticed superimposed on the high-frequency oscillations typical of high-order singular functions. Since the same does not occur in the singular functions of corresponding order for the single-step operator $S_N$ (as shown in figure 8), it is reasonable to interpret the singular functions of the flat region of the spectrum of $D_N$ as the functions generating the null functions for $D$ (in section 4.1 it is pointed out that these null functions are indeed characterized by low-frequency oscillations).

**Remark 4.1.** The distinctive feature of the discretized double-step kernel (non-uniform ionization) inverse problem is that it has high-order (low $\sigma_k$) singular functions containing low-frequency components. These correspond to the low-frequency structure imposed on the photon spectrum by the plasma step ionization structure and have the effect of filtering out information on the electron spectrum at energies related to the transition step depth. This is equivalent to low-frequency non-uniqueness in the continuous problem and contrasts with the usual high-frequency filtering characteristic of ill-posed inverse problems.

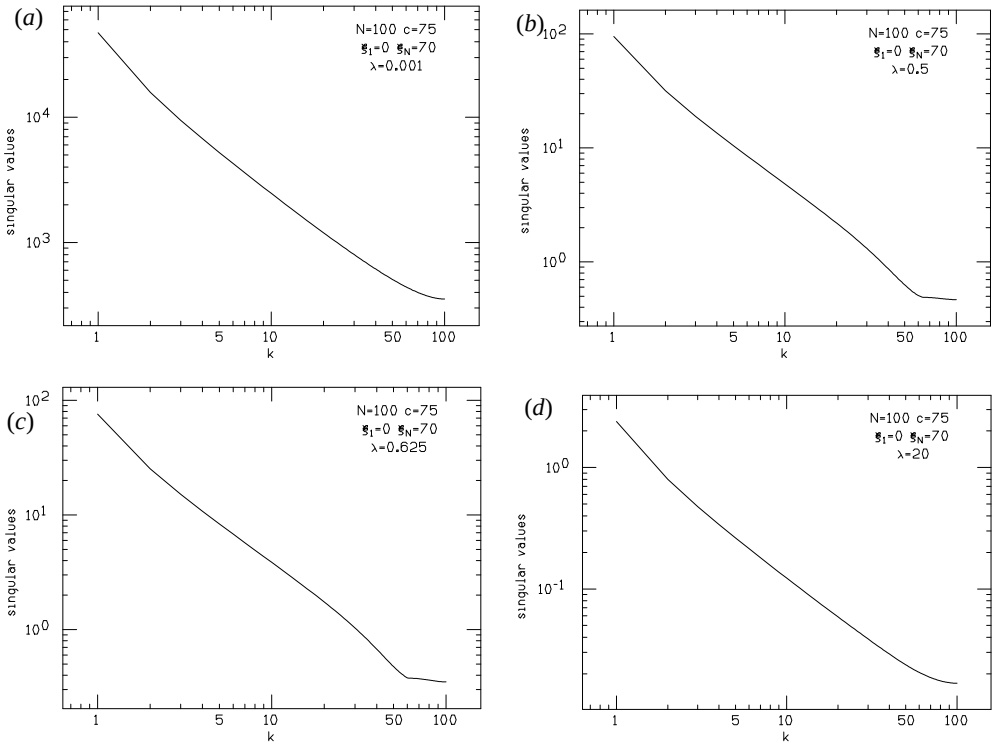


**Figure 5.** Singular spectrum of $D_N$ for different values of $\lambda$. (*a*) $\lambda = 0.001$; (*b*) $\lambda = 0.5$; (*c*) $\lambda = 0.625$; (*d*) $\lambda = 20$. $N = 100$ points are uniformly sampled in the range $\xi_1 = 0$, $\xi_N = 70$ with $c = 75$.

## 5. Solution of the inverse problem

The availability of the singular system for the finite-rank operators $S_N$ and $D_N$ significantly eases the solution of the problem of determining the electron distribution function $f$ from a photon spectrum data set $\boldsymbol{g}$ both for uniform and non-uniform ionization conditions. We briefly describe this fact in the case of $D_N$.

Let us consider the least-squares problem

$$\|\boldsymbol{g} - D_N f\|_Y = \text{minimum} \tag{5.1}$$

with $\boldsymbol{g} = (g_1, \ldots, g_N)$ in $Y$ and $\|\cdot\|_Y$ the norm induced by $(\cdot,\cdot)_Y$. This minimum problem is equivalent to the Euler equation

$$D_N^* D_N f = D_N^* \boldsymbol{g} \tag{5.2}$$

and the set of its solutions is closed and convex. It follows that there exists a unique minimum norm least-squares solution $f^\dagger$ which is called the generalized solution of problem (4.13) and which can be described in terms of the singular system by [8]

$$f^\dagger(\eta) = \sum_{k=1}^{N} \frac{1}{\sigma_k^{(N)}} (\boldsymbol{g}, \boldsymbol{v}_k^{(N)})_Y u_k^{(N)}(\eta). \tag{5.3}$$

Therefore, a general representation for the solutions of (4.13) is

$$f = f^\dagger + h \tag{5.4}$$

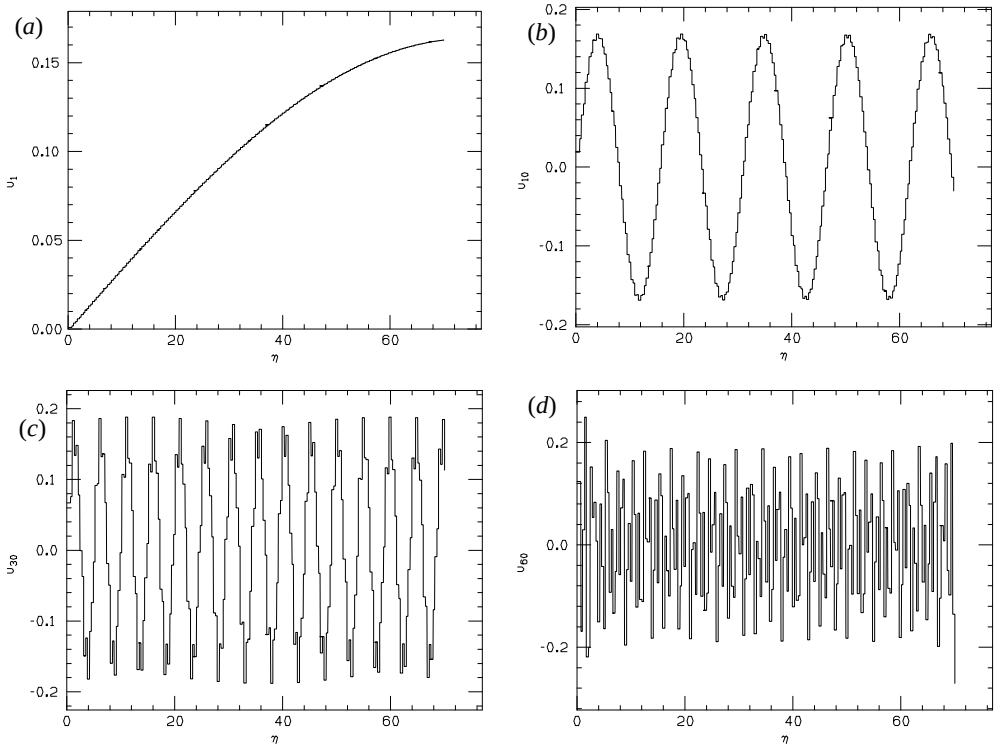


**Figure 6.** Singular functions $u_k$ of $D_N$ in the case $\xi_1 = 0$, $\xi_N = 70$, $N = 100$, $c = 75$ and $\lambda = 0.625$. (*a*) $k = 1$; (*b*) $k = 10$; (*c*) $k = 30$; (*d*) $k = 60$.

with $h$ in the kernel of $D_N$. Owing to the presence of small singular values the generalized solution (5.3) is not physically significant when the data vector $\boldsymbol{g}$ is affected by the noise consequence of the measurement procedure. This is the reason why regularization methods must be introduced to provide stable approximate solutions of the inverse problem. Also, for some significant algorithms the regularized solutions can be written in terms of the singular system. This is the case, for example, of the truncated singular value decomposition (TSVD) where the regularized solution is obtained by stopping the sum (5.3) at some number $M < N$. Another example is the Tikhonov method where the regularized solution is chosen in the one-parameter family of solutions of the minimum problem

$$\|D_N f_\alpha - \boldsymbol{g}\|_Y^2 + \alpha \|f_\alpha\|_{L^2(0,c)}^2 = \text{minimum} \tag{5.5}$$

or, equivalently, of the Euler equation

$$(D_N^* D_N + \alpha I) f_\alpha = D_N^* \boldsymbol{g}. \tag{5.6}$$

It is not difficult to prove the following expansion for $f_\alpha$:

$$f_\alpha(\eta) = \sum_{k=1}^{N} \frac{\sigma_k}{\sigma_k^2 + \alpha} (\boldsymbol{g}, \boldsymbol{v}_k)_Y u_k(\eta). \tag{5.7}$$

Both in the TSVD and in the Tikhonov method the optimal choice of the regularization parameters $M$ and $\alpha$ is performed by using *ad hoc* criteria whose computation is again eased by using the singular system of the operator. It is our intention to apply these regularization methods (and also other ones which explicitly allow for *a priori* knowledge of the solution) to the inversion of the real data the HESSI mission will provide next year.

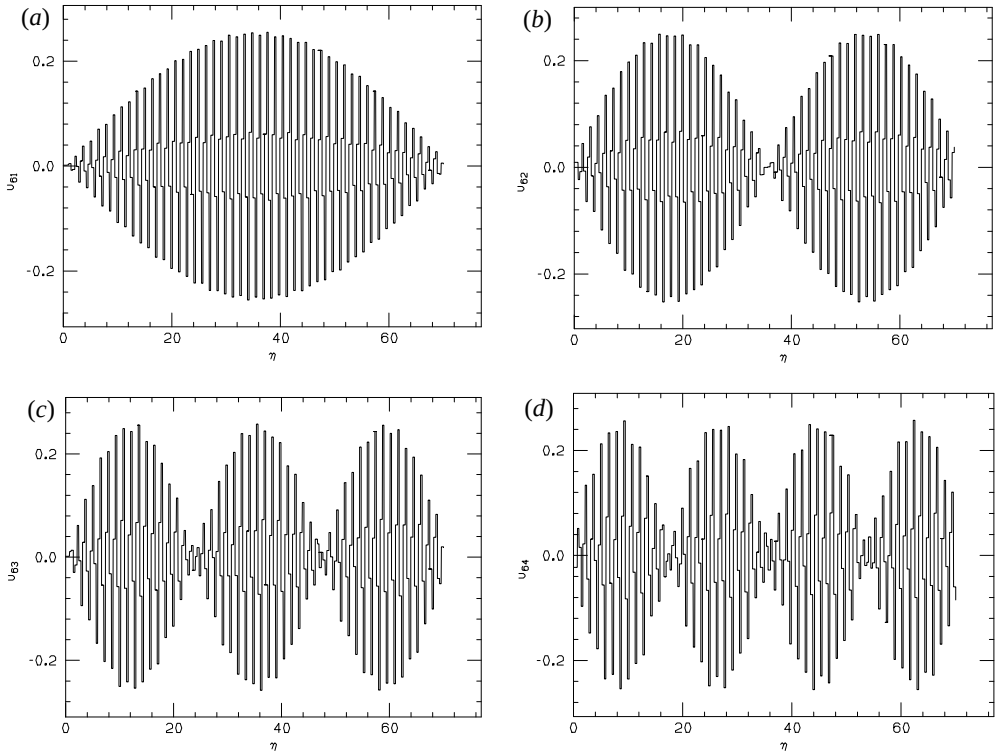


**Figure 7.** Singular functions $u_k$ of $D_N$ in the case $\xi_1 = 0$, $\xi_N = 70$, $N = 100$, $c = 75$ and $\lambda = 0.625$. (*a*) $k = 61$; (*b*) $k = 62$; (*c*) $k = 63$; (*d*) $k = 64$.

## Appendix. Equations for general $q(\epsilon, E)$

Here we show that the problem of non-uniqueness does not arise because of the form of cross section $q$ adopted. If we reverse the order of integration in (2.1) it can be written as a two-stage operation, namely

$$\mathcal{Q}(\mathcal{F}(E);\epsilon) = \int_{\epsilon}^{\infty} q(\epsilon, E)\mathcal{F}(E)\,\mathrm{d}E = H(\epsilon) \tag{A.1}$$

with

$$\int_{E}^{\infty} \frac{F_0(E_0)\,\mathrm{d}E_0}{\lambda + x(E_0, E)} = \mathcal{F}(E_0). \tag{A.2}$$

Solution of the first equation for $\mathcal{F}$ amounts to deconvolution of a mean source electron spectrum through the cross section operator $q$, formally

$$\mathcal{F}(E) = \mathcal{Q}^{-1}[H(\epsilon); E] \tag{A.3}$$

while solution of the second amounts to deconvolution of the resulting $\mathcal{F}$ over the electron path, allowing for the plasma ionization structure. Although the form of $\mathcal{F}$ will depend on the $q$ used, the solution (A.3) is unique, but when $\mathcal{F}$ is inserted in (A.2) the problem of non-uniqueness for a step function $x(N)$ remains. An important further issue which we do not address here is whether the non-uniqueness arises for more general forms of $x$ than a step function. We merely observe here that, for general $q$, $K$, differentiation of (A.2) with respect to $E$ yields

$$F_0(E_0) - \frac{1}{K(0)} \int_{E}^{\infty} K'(E_0^2 - E^2) F_0(E_0)\,\mathrm{d}E_0 = -\frac{\mathcal{F}'(E_0)}{K(0)}. \tag{A.4}$$

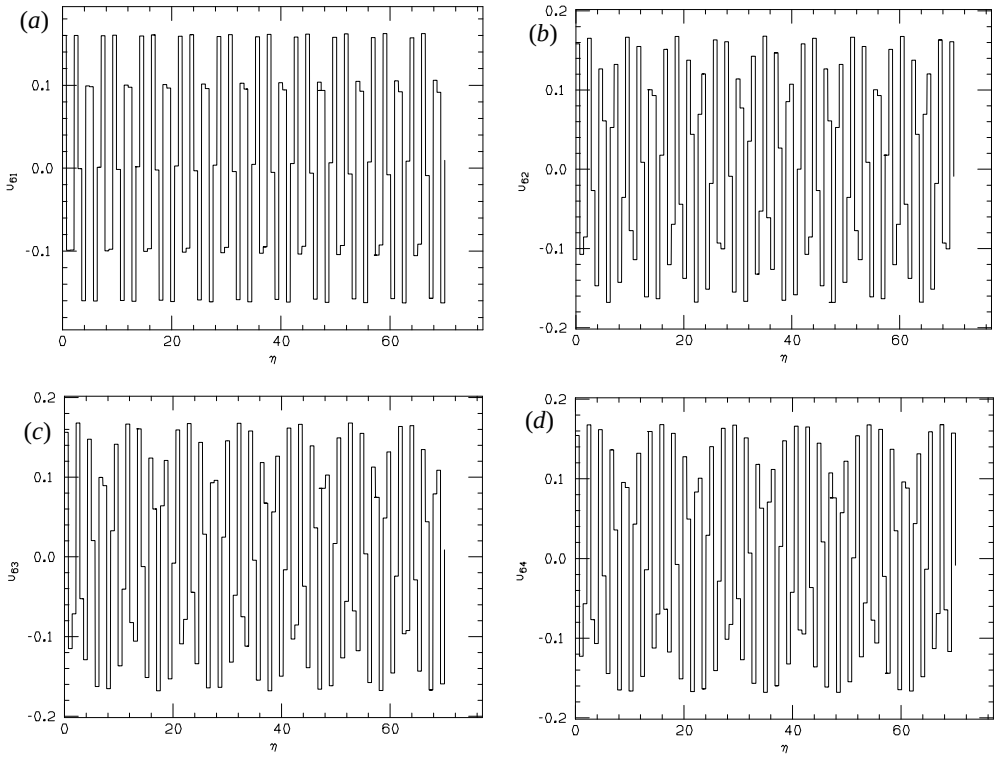


**Figure 8.** Singular functions $u_k$ of $S_N$ in the case $\xi_1 = 0$, $\xi_N = 70$, $N = 100$ and $c = 75$. (*a*) $k = 61$; (*b*) $k = 62$; (*c*) $k = 63$; (*d*) $k = 64$.

In the case of constant $x$, $K' = 0$ and the integral term vanishes. In the case of a step function $x$, $K'$ is a delta function and the integral is replaced by a shift operator as in (2.9). What is required is examination of whether the general (non-step) form of $K$ leads to the operator in (A.2) having null functions. We do not pursue this issue here but simply make some observations. For the step function ionization structure it is clear that there is a set of exactly zero singular values corresponding to singular functions with long wavelengths. It seems quite plausible that the existence of these singular functions is a result of the particular (discontinuous) form of the kernel, and that they will disappear for other, more realistic, ionization structures. However, if we consider a sequence of inverse problems with increasingly sharp, but continuous, kernels what we expect to see is the existence of singular values corresponding to long-wavelength singular functions that are non-zero in general but which tend to zero as the kernel approaches a step function. So, for a realistic ionization structure with a sharp but continuous change, we expect to see a set of *very small* singular values corresponding to long-wavelength singular functions. In practical applications with noisy data such distinction between 'small' and 'zero' singular values is, of course, irrelevant.

## References

[1] Bertero M, De Mol C and Pike E R 1985 Linear inverse problems with discrete data: I. General formulation and singular system analysis *Inverse Problems* **1** 301

[2] Bertero M and Grünbaum F A 1985 Commuting differential operator for the finite Laplace transform *Inverse Problems* **1** 181–92

[3] Brown J C 1971 The deduction of energy spectra of non-thermal electrons in flares from the observed dynamic spectra of hard x-ray bursts *Sol. Phys.* **18** 489
[4] Brown J C, McArthur G K, Barrett R K, McIntosh S W and Emslie A G 1998 Inversion of thick-target bremsstrahlung spectra from nonuniformly ionized plasma *Sol. Phys.* **179** 379–404
[5] Brown J C and Craig I J D 1986 *Inverse Problems in Astronomy* (Bristol: Hilger)
[6] Dennis B R *et al* 1996 High energy solar spectroscopic imager (HESSI) *Proc. SPIE* **2804** 228
[7] Johns C M and Lin R P 1992 The derivation of parent electron spectra from bremsstrahlung hard x-ray spectra *Sol. Phys.* **137** 121
[8] Kato T 1980 *Perturbation Theory for Linear Operators* (Berlin: Springer)
[9] Koch H W and Motz J W 1959 Cross-section formulae and related data *Rev. Mod. Phys.* **31** 920
[10] Lin R P and Schwartz R A 1987 High spectral resolution measurements of a solar flare hard x-ray burst *Astrophys. J.* **312** 462
[11] Parker E N 1994 *Spontaneous Current Sheets in Magnetic Fields with Applications to Stellar X-Rays* (Oxford: Oxford University Press)
[12] Piana M 1994 Inversion of bremsstrahlung spectra emitted by solar plasma *Astron. Astrophys.* **288** 949–59
[13] Piana M and Brown J C 1998 Optimal inversion of hard x-ray bremsstrahlung spectra *Astron. Astrophys. Suppl.* **132** 291–9
[14] Priest E R 1984 *Solar Magnetohydrodynamics* (Dordrecht: Reidel)
[15] Sturrock P A 1980 *Solar Flares: a Monograph From Skylab Solar Workshop II* (Boulder, CO: Colorado Associated University Press)
[16] Tandberg-Hanssen E and Emslie A G 1988 *The Physics of Solar Flares* (New York: Cambridge University Press)
[17] Thompson A M, Brown J C, Craig I J D and Fulber 1992 Inference of non-thermal electron energy distributions from hard x-ray spectra *Astron. Astrophys.* **265** 278